Molecular Diagnosis of Inherited Retinal Diseases in Indigenous African Populations by Whole-Exome Sequencing

Lisa Roberts,[1] Rinki Ratnapriya,[2] Morné du Plessis,[1] Vijender Chaitankar,[2] Raj S. Ramesar,[1] and Anand Swaroop[2]

[1]University of Cape Town/MRC Human Genetics Research Unit, Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[2]Neurobiology, Neurodegeneration & Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, Maryland, United States

Correspondence: Anand Swaroop, NNRL, National Eye Institute, MSC0610, 6 Center Street, Bethesda, MD 20892, USA; swaroopa@nei.nih.gov.
Lisa Roberts, Room N3.14, Level 3, Wernher and Beit North Building, Institute of Infectious Disease and Molecular Medicine, University of Cape Town Faculty of Health Sciences, Anzio Road, Observatory, 7925, Cape Town, Western Cape, South Africa; lisa.roberts@uct.ac.za.

LR and RR contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: April 21, 2016
Accepted: October 18, 2016

Citation: Roberts L, Ratnapriya R, du Plessis M, Chaitankar V, Ramesar RS, Swaroop A. Molecular diagnosis of inherited retinal diseases in indigenous African populations by whole-exome sequencing. *Invest Ophthalmol Vis Sci.* 2016;57:6374–6381. DOI: 10.1167/iovs.16-19785

Purpose. A majority of genes associated with inherited retinal diseases (IRDs) have been identified in patients of European origin. Indigenous African populations exhibit rich genomic diversity, and evaluation of reported genetic mutations has yielded low returns so far. Our goal was to perform whole-exome sequencing (WES) to examine variants in known IRD genes in underrepresented African cohorts.

Methods. Whole-exome sequencing was performed on 56 samples from 16 families with diverse IRD phenotypes that had remained undiagnosed after screening for known mutations using genotyping-based microarrays (Asper Ophthalmics). Variants in reported IRD genes were identified using WES and validated by Sanger sequencing. Custom TaqMan assays were used to screen for identified mutations in 193 unrelated indigenous Africans with IRDs.

Results. A total of 3494 variants were identified in 217 known IRD genes, leading to the identification of seven different mutations (including six novel) in six genes (*RHO*, *PRPF3*, *PRPF31*, *ABCA4*, *CERKL*, and *PDE6B*) in six distinct families. TaqMan screening in additional probands revealed identical homozygous *CERKL* and *PDE6B* variants in four more patients.

Conclusions. This is the first report of WES of patients with IRDs in indigenous African populations. Our study identified genetic defects in almost 40% of the families analyzed, significantly enhancing the molecular diagnosis of IRD in South Africa. Thus, WES of understudied cohorts seems to present an effective strategy for determining novel mutations in heterogeneous retinal diseases.

Keywords: next generation sequencing, genetic testing, photoreceptor dysfunction, South Africa, vision loss, inherited blindness, retinal degeneration, clinical genetics

Indigenous African populations are underrepresented in international genetic/genomic studies. The African continent includes 55 countries (https://africacheck.org/reports/how-many-countries-in-africa-how-hard-can-the-question-be/), with over 2000 distinct ethnolinguistic groups.[1] Being the most ancient of all populations, Africans display vast genetic diversity[2,3] as a result of historical migration, population admixture, response to environmental change, and/or exposure to a plethora of infectious agents.[1] Indigenous Bantu language–speaking individuals arrived in South Africa approximately 1500 years ago as a result of the movement of people, known as the "Bantu expansion," across (west to east) and down (north to south) Africa.[4,5] Subsequent divergence of Bantu speakers in South Africa occurred relatively recently into separate ethnolinguistic groups such as Sotho-Tswana, Xhosa, and Zulu. These black South African individuals, referred to collectively hereafter as indigenous Africans, are the focus of this study as they provide a valuable resource to detect genetic defects in heterogeneous Mendelian diseases including inherited retinal diseases (IRDs).

Inherited retinal diseases encompass a genetically and clinically heterogeneous group of blinding diseases, with a common phenotype of dysfunction and/or degeneration of the light-sensitive photoreceptor cells (rods and cones) in the retina.[6,7] Patients with gene defects causing a primary disease of rod photoreceptors, for example, retinitis pigmentosa (RP), initially experience night blindness and loss of peripheral vision. In contrast, IRDs showing initially the loss of cone photoreceptors, for example, macular degeneration (MD) and Stargardt disease (STGD), manifest with a loss of central vision. Inherited retinal diseases can exhibit autosomal dominant, autosomal recessive, or an X-linked pattern of inheritance and demonstrate progressive or stationary and syndromic or non-syndromic clinical phenotypes.[6,7] Over 240 genes have been

identified for IRDs (https://sph.uth.edu/Retnet/sum-dis.htm; in the public domain). Recent studies using animal models have finally begun to uncover some of the underlying disease mechanisms and pathways that affect photoreceptor development or function.[7–9] Furthermore, it is estimated that only 50% to 70% of the cases with RP (depending on geographical regions or populations) can be attributed to the known genes,[10–12] indicating that a considerable number of as yet unknown mutations and genes remain to be identified. Such a vast clinical and genetic heterogeneity displayed by IRDs confounds molecular diagnosis and investigation of the pathogenic mechanisms.

Identification of the specific genetic defect in a patient with IRD affords several potential benefits. First, overlapping phenotypes and clinical variability of IRDs do not always permit a clear clinical (ophthalmologic) diagnosis/prognosis. Genetic analysis is unequivocal and provides clinical utility as diagnostic, predictive, and carrier testing can be offered to family members. Second, genetic tests may also influence the clinical management of the disease. The IRD research program in South Africa (SA), initiated in 1990,[13] has a strong translational and service component.[14] Lastly, the knowledge of precise genetic defects can allow development of gene-based therapies for treatment of IRDs.[15]

The reported prevalence of IRDs is approximately 1 in 3500[11] in populations where epidemiologic data are available. No data exist on the prevalence of IRDs in Africa. Nonetheless, using SA's 2011 population census (http://www.statssa.gov.za/; in the public domain), one may extrapolate that approximately 14,500 individuals suffer from IRD-related visual impairment/blindness in SA; of these (taking population demographics into account), as many as 11,600 are expected in the indigenous African population. However, a high frequency of unaffected carriers of IRD gene mutations could exist because of local founder effects and further elevate the potential burden of disease.[11]

Demographic information, biological material, clinical details, and diagnoses have been archived for 3237 individuals in 1430 SA families with distinct IRDs in the University of Cape Town (UCT) registry, which contains information and biological material primarily from individuals of Caucasian origin; indigenous Africans currently comprise only 19% of the collection ($n = 275$ families). Understandably, this does not reflect the population demographics of SA and is due to ascertainment bias and the lack of resources in rural areas where a large proportion of the indigenous populations reside. To date, 249 families (249/1430 = 17%), mostly Caucasian ($n = 204/249$; 82%), are in diagnostic mode, with clear pathogenic mutations having been identified using a variety of methods.[16] The most prevalent reported genetic defects in IRDs exhibit an almost insignificant incidence in the SA patient cohort.[17–20] Investigation of the indigenous African subcohort for reported mutations through the use of Asper Ophthalmics microarrays (http://www.asperbio.com/asper-ophthalmics; in the public domain) has produced lower returns in the indigenous African IRD subcohort than in the Caucasian subcohort. Approximately 41.2% of Caucasian samples ($n = 279$) have been diagnosed by microarray screening as opposed to only 12.8% of indigenous African samples ($n = 109$) because each Asper Ophthalmics microarray specifically tests for reported mutations that have been identified predominantly in patients of European/Caucasian origin. Novel mutations are detected only if they occur at a nucleotide position(s) where a mutation has already been reported, as only select nucleotides are assayed. Thus, either SA indigenous IRD patients harbor novel mutations in known genes that are not included in the Asper arrays or causative genes are novel.

The advent of next-generation sequencing technologies has revolutionized the speed and cost at which disease mutations can be identified. An increased number of mutations are now being identified in different populations using high-throughput methods such as whole-exome sequencing (WES).[21,22] Improved molecular diagnosis in patients is important, given the number of clinical trials and treatments currently under investigation for this group of disorders.[23] We therefore resorted to a comprehensive WES approach, followed by targeted analysis of all reported IRD genes, toward understanding the genetic architecture of IRD in the indigenous SA population.

Materials and Methods

Patient Cohort

Informed consent was obtained according to the 2008 Declaration of Helsinki for all members from whom samples have been archived in the UCT IRD registry. Ethics approval was granted by the Human Research Ethics Committee of the UCT Faculty of Health Sciences (Rec Ref. 226/2010 and 768/2013). Samples from indigenous African families were selected from the registry if DNA was available from at least three family members and if a proband had been screened using the appropriate microarray but no molecular diagnosis had been obtained. A total of 16 families met the selection criteria, comprising 109 individuals; of these, 56 were chosen for WES. The selected 16 families originated from diverse, self-identified, indigenous African ethnolinguistic groups: 5 Xhosa, 3 Zulu, 2 Tswana, 1 Shangaan, 1 Venda, 1 Tsonga/Ndebele, 1 Xhosa/Sotho, and 2 Unknown. Two of the 16 families had been clinically diagnosed with autosomal recessive MD (one of whom had a subsequent diagnosis of Leber congenital amaurosis) and 14 with RP.

Whole-Exome Sequencing

Genomic DNA samples were quantified using the QuantiFluor dsDNA system (Promega, Madison, WI, USA), according to manufacturer's instructions. Whole-exome capture was performed on 50 ng DNA using the Nextera Rapid Capture Expanded Exome kit (Illumina, San Diego, CA, USA), and 125-bp paired-end sequences were obtained on a HiSeq2500 platform (Illumina), according to manufacturer's instructions. Details of WES analysis are described elsewhere.[24] FastQC (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; in the public domain) was used to confirm quality of sequencing, after which adapter indexes were removed using Trimmomatic.[25] Reads were mapped to the human reference sequence (hg19, GRCH37) using BWA,[26] and GATK[27,28] was used for variant calling, local realignment, base quality recalibration, and variant recalibration. Annotation of variants was performed with ANNOVAR.[29]

Variant Prioritization and Validation

Sequence variants present in genes (Supplementary Table S1) listed on the RetNet database (https://sph.uth.edu/Retnet/sum-dis.htm; in the public domain; accessed 12 November 2014) were extracted for further analysis. Variants with a minor allele frequency (MAF) of <0.1 in the 1000 Genomes Project[30] (October 2014 annotation) were prioritized, as were exonic or splicing variants. The variants were subsequently selected based on cosegregation with the disease phenotype within each family. For nonsynonymous variants, a minimum threshold of three pathogenic predictions was applied to the dbNSFP[31] annotation of ANNOVAR, for either of the following

Table 1. Candidate Variants in Each of the 16 Families After Prioritization Filters

Family ID	No. of IRD Variants	<0.1 MAF	Exonic/Splicing	Cosegregating Within Family	Pathogenic, >3 Predictions	Candidate Gene, Rare and Cosegregating
RPD 55	1351	749	280	17	7	0
RP 583	1431	796	302	8	8	0
RPD 94	1181	599	198	10	1	0
RP 391	1224	607	209	25	13	*PRPF3*
RPD 401	1183	619	234	30	11	0
RPD 799	1309	686	259	15	5	0
RPD 1001	1416	805	316	8	4	0
RPD 1005	1285	679	223	5	3	0
RPD 1010	1217	628	234	5	3	*RHO*
RPD 1339	1153	579	194	21	10	*PRPF31*
RPM 537	1130	550	191	9	5	*ABCA4* (x2)
RPM 1167	1086	552	198	2	0	0
RPR 397	1063	525	199	19	3	*PDE6B*
RPR 624	1200	620	217	3	0	0
RPR 917	1154	574	203	4	1	*CERKL*
RPX 54	1432	760	259	1	1	0

predictor subsets: (SIFT, PolyPhen2-HDIV, PolyPhen2-HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, MetaSVM, and MetaLR), or (VEST3 CADD-raw, CADD-phred, GERP++, phyloP46way-placental, phyloP100way-vertebrate and SiPhy-29way-logOdds). Variants were then assessed for their presence in the remainder of the cohort. High-priority candidate variants were finally evaluated by examining RetNet and Ensembl release 8[32] with particular emphasis on population data for 1000 Genomes African subpopulations and NHLBI Exome Sequencing data (http://evs.gs.washington.edu/EVS/; in the public domain) in African Americans, as well as reported phenotypes associated with the genes.

Wherever possible, additional familial samples not subjected to WES were included for validation of candidate variants by Sanger Sequencing on a 3130xl Genetic Analyzer (Applied Biosystems, ThermoFisher Scientific, Waltham, MA, USA). Finally, validated variants were checked for MAF in the African Genomes Variation Project[33] data, which include low-coverage whole-genome sequences from 100 Bagandan of Uganda, 100 Zulu of SA, and 120 Ethiopian individuals.

Screening of SA Cohort

Custom TaqMan assays (primer and reporter sequences in Supplementary Table S2) were designed to determine the allele frequency of seven variants identified by WES, in a larger cohort of 193 unrelated indigenous African probands with IRDs but no known causative mutation. In order to determine the optimal template concentration, two control samples were screened for each assay (including a positive control for each), at 8, 6, 4, 3, 2, and 1.5 ng/μL. It was empirically determined that 2 ng/μL was optimal, allowing for effective allele discrimination for each assay.

The final volume in each assay reaction was 5 μL, composed of 2.5 μL TaqMan GT mastermix (2×) (Applied Biosystems), 0.25 μL assay mix (20×), 2.25 μL DNA (at 2 ng/μL, that is, total input of 4.5-ng template). Each assay included at least two no-template controls and two positive controls. Thermal cycling was performed using the ABI 7900HT instrument (Applied Biosystems) and the following conditions: 95°C, 10 minutes; (95°C, 15 seconds; 60°C, 1 minute) × 40 cycles. If fluorescence values dictated after this cycling, a second cycling of 10× (95°C, 15 seconds; 60°C, 1 minute) cycles and subsequent postread analysis were performed. Sanger sequencing was used to validate all candidate variants.

Results

Whole-exome sequencing was performed for 56 samples that included at least three individuals from each of the 16 families. On average, 92% of the exome was captured at 25× coverage, and a total of 1,816,031 variants were identified. We excluded intergenic (n = 759,459), intronic (n = 710,303), and synonymous (n = 59,723) variants from further analysis and identified 3494 candidate variants in 217 reported IRD genes. We then filtered out variants that were present upstream or downstream (n = 298) of the coding exons, in the 5′ or 3′ untranslated region (n = 1813), or in the noncoding RNA (ncRNA) regions (n = 96). Of the remaining IRD variants (1266 exonic and 21 splice site), 561 variants were potentially pathogenic (Supplementary Table S3). At least three prediction algorithms identified 498 variants as pathogenic, and 63 variants were deletions, insertions, gain/loss of stop codons, or variants of unknown effect. The candidate variants remaining after each filtering step are shown in Table 1.

We identified seven different likely mutations in six IRD families; of these, six had not been reported previously (Table 2; Fig. 1). Four of the variants are missense, one is predicted to affect splicing, and two are predicted to result in frameshift and protein truncation. None of the variants has been reported in the whole-genome sequence data of 100 Zulu, 100 Bagandan, or 120 Ethiopian individuals in the AGVP study.[33] Additionally, these variants are not detected in 97 Luyha or 88 Yoruba individuals in the 1000 Genomes data.[30] Therefore, the seven variants identified in IRD families are not present in 505 control African individuals (1010 chromosomes), providing additional evidence in support of their pathogenicity. The previously reported autosomal recessive RP (arRP) mutation p.(His620GlnfsTer23) in *PDE6B* was present only once in 4266 alleles in the NHLBI WES dataset (ESP) of African Americans (rs769671323, as of 27 October 2015); this frameshift mutation is predicted to generate a truncated protein lacking over 200 C-terminal amino acids.[34] The second frameshift mutation identified in *ABCA4* is predicted to truncate the protein by 612 C-terminal amino acids. The c.698-1G>A variant in the acceptor splice site of exon 8 of *PRPF31*, interrogated by Human Splicing Finder 3.0,[35] is predicted to activate an intronic cryptic acceptor site while simultaneously disrupting an exon splicing silencer site and creating an exon splicing enhancer site. Therefore, all seven variants were computationally predicted to be pathogenic, cosegregated with disease in

Table 2. Potential Causative Mutations in Indigenous African Families With IRDs

Family	Disorder	Ethnicity	Gene	Variant: cDNA; Protein	Comment	Pathogenicity, ACMG Category[36]	Reported/Novel
RP 391	adRP	Tswana	*PRPF3*	c.1480A>G; p.(Thr494Ala)	Heterozygous, 9 pathogenic predictions	Likely pathogenic	Novel
RPD 1010	adRP	Xhosa	*RHO*	c.154T>G; p.(Phe52Val)	Heterozygous, 4 pathogenic predictions	Likely pathogenic	Novel
RPD 1339	adRP	Zulu	*PRPF31*	c.698-1G>A; p.(?)	Heterozygous	Likely pathogenic	Novel
RPM 537	arSTGD	Venda	*ABCA4*	c.4832delC; p.(Thr1611MetfsTer51) c.1043T>G; p.(Phe348Cys)	Compound heterozygous, 9 pathogenic predictions for nonsynonymous p.(Phe348Cys)	Pathogenic (frameshift truncation), likely pathogenic (missense)	Both novel
RPR 397	arRP	Shangaan	*PDE6B*	c.1860delC; p.(His620GlnfsTer23)	Homozygous	Pathogenic	Reported, Danciger et al.[34]
RPR 917	arRP	Xhosa	*CERKL*	c.365T>G; p.(Leu122Arg)	Homozygous, 4 pathogenic predictions	Likely pathogenic	Novel

adRP, autosomal dominant RP; arRP, autosomal recessive RP; arSTGD, autosomal recessive STGD.

the respective families, verified by Sanger sequencing, and exhibited conservation across vertebrates (Fig. 2). According to ACMG guidelines for the interpretation of sequence variants,[36] the frameshift truncations identified in this study have sufficient evidence to classify them as "pathogenic," while each of the splice site or missense variants meets the criteria of "likely pathogenic" variants in the absence of functional studies.

We then performed TaqMan assays for these seven pathogenic or likely pathogenic variants, identified here, in an additional 193 indigenous Africans with IRDs. Five of these variants were not detected in this cohort. The *PDE6B* c.1860delC mutation was identified in a homozygous state in one additional individual (diagnosed with arRP, from infancy) and in a heterozygous state in four individuals (two sporadic RP, one arRP, and one with an apparent dominant family history). In addition, we identified the homozygous *CERKL* c.365T>G variant in three patients with different IRD phenotypes: one each of sporadic RP, sporadic STGD, and arRP. This c.365T>G variant was also identified in the heterozygous state in one RP proband.

Discussion

The use of indigenous SA populations, combined with next-generation sequencing platforms, provides an enriched resource for discovering novel IRD genes and mutations. Due to the vast clinical and genetic heterogeneity, traditional candidate gene-based approaches have been less effective for the molecular diagnosis of IRDs. Targeted capture of specific IRD genes, associated with particular retinal phenotypes, is a strategy being used for molecular diagnosis with increasing frequency.[37–41] Both targeted capture and WES allow for the detection of novel mutations in genes (in contrast to microarrays). Recently targeted capture of known IRD genes in panel-based testing was reportedly more successful than WES,[42] probably due to better coverage of the genes of interest. We believe that panel-based testing is especially not suitable for the research on understudied populations, like the indigenous Africans, where WES with targeted bioinformatic analysis could enhance molecular diagnosis and even lead to novel gene discovery. Collaborative and combined analysis of WES data from different groups can yield genetic evidence for novel IRD genes. In addition, WES data from unresolved families can be reanalyzed when novel IRD genes are reported without redesigning diagnostic gene panels and performing new experiments. The latter is an important consideration when providing a molecular diagnosis for patients in resource-limited settings.

Our targeted analysis approach was successful in assigning molecular diagnosis in 38% of the indigenous African families, a clear improvement on the 13% detection rate using the commercially available arrays that test for specific reported variants. Six of seven (85%) variants discovered were novel, supporting the high genetic heterogeneity in IRDs as well as genetic diversity among indigenous Africans. Analysis of a larger cohort of unrelated indigenous African probands revealed that five out of seven variants were rare and detected in a single family each, further advocating the use of WES-based diagnosis instead of the genotyping-based microarrays used previously to screen this population group. Nonetheless, the detection rate is still much lower than the reported 83% of European families interrogated using a similar approach.[43] Other population groups investigated in a comparable manner include Saudi Arabian,[44] Chinese,[45] Thai,[46] and Israeli,[22,47] with detection rates ranging from 49% to 83% and the number of analyzed genes ranging from 60 to 226.

The relatively low detection of causal mutations in the SA cohort of IRD families can be attributed to multiple factors. Whole-exome sequencing is a capture-based method with genomic regions of low coverage and poor detection of large genomic alterations. Additionally, WES will not detect less obvious pathogenic variants, such as ncRNA or regulatory variants and those present in the untranslated regions or introns. The clinical complexity of IRDs, that is, nonpenetrance, frequent manifestation in carriers of X-linked disorders, and variable expressivity within families, could result in an incorrect inheritance pattern being assumed and hence incorrect variant filtering during cosegregation analysis. This problem is exacerbated in SA, where frequently sparse clinical information accompanies the samples particularly from the more rural areas of the country, and where language barriers can often result in misinformation. However, it is also plausible that causative mutations in many families reside in an as yet

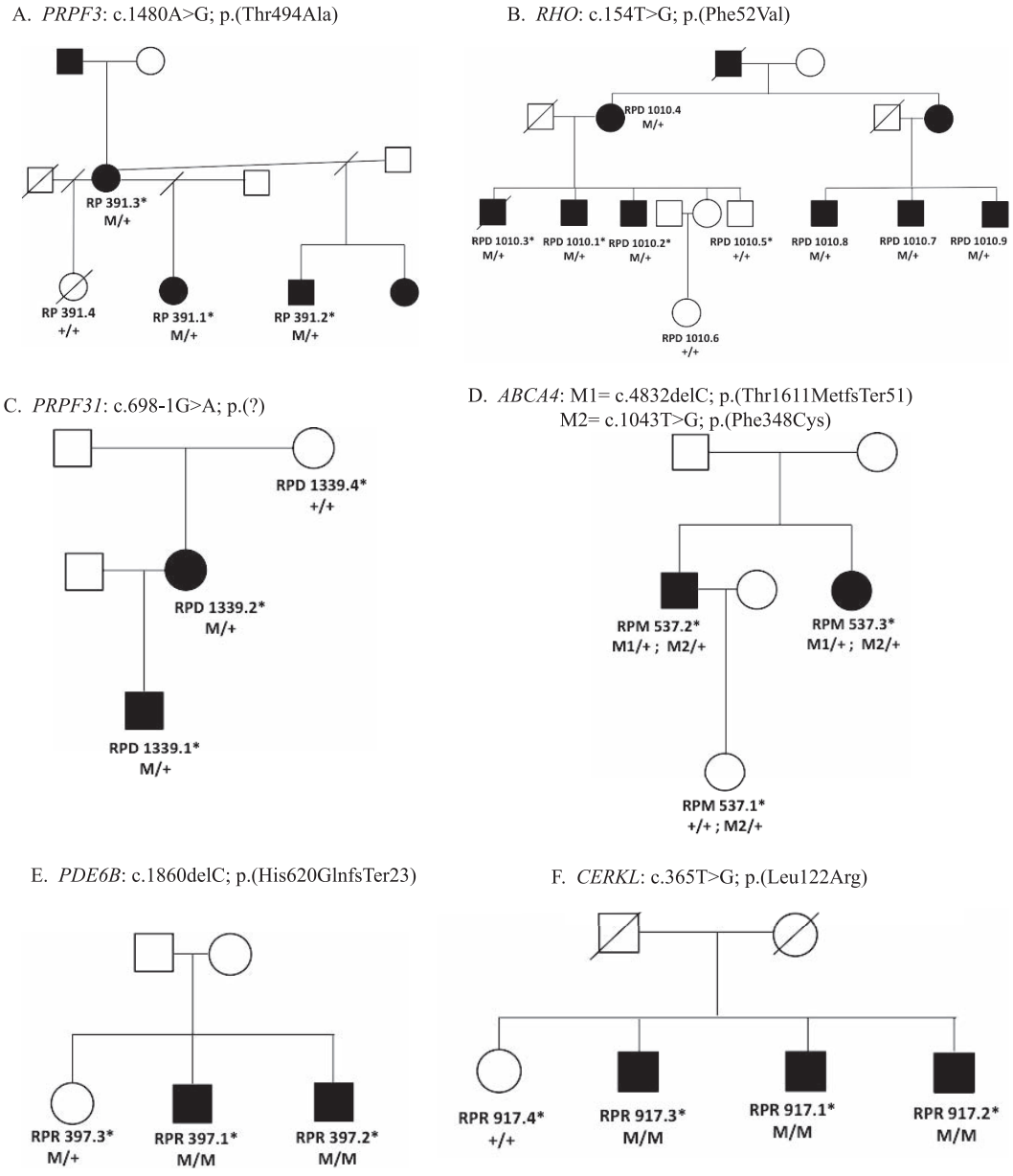

Figure 1. Pedigrees of IRD families showing cosegregation of the variants identified by WES. *Squares* represent males, and *circles*, females. *Shaded symbols* indicate individuals with IRD. Identifier codes show individuals from whom biological material is available, and those selected for whole-exome sequencing are noted with an *asterisk*. Segregation of mutation(s) in the families is indicated as +/+, homozygous for wild-type allele; M/+, heterozygous; M/M, homozygous for mutation. Clinical information is presented in Supplementary Table S4.

unreported IRD gene. We believe that the use of previously understudied populations is a sensible approach for ascertaining missing heritability in genetically heterogeneous diseases such as IRDs.

PDE6B mutations have been associated with autosomal dominant congenital stationary night blindness (adCSNB) and arRP.[48] In our patient samples, two probands with arRP carried the homozygous c.1860delC mutation of *PDE6B*. In addition, we identified four IRD patients (two sporadic RP, one arRP, and one with an apparent autosomal dominant [adRP] family history) with a heterozygous *PDE6B* c.1860delC allele. The relatively high frequency of this allele (1.9%; $n = 8/418$ alleles) in the SA IRD cohort could imply compound heterozygosity for *PDE6B*, digenic inheritance, or enhanced genetic burden. The individual RPR 397.1 (in the WES cohort) had been tested previously by the arRP microarray; however, this array platform was designed to detect the c.1857_1858delC *PDE6B* variant and not c.1860. We also noted the relatively frequent occurrence of the *CERKL* c.365T>G variant in SA IRD patients ($n = 9/418$ alleles; 2.2%). The four homozygous cases with this mutation displayed varying phenotypes: two arRP, one sporadic RP, and one sporadic STGD. *CERKL* mutations are shown to result in autosomal recessive forms of cone dystrophy, cone–rod dystrophy and RP (RetNet). In our study, an identical *CERKL* mutation is associated with distinct IRD phenotypes, implying the existence of modifier variants or the impact of vastly different environmental and epigenetic landscape in this genetically diverse cohort compared to the reported Caucasian patients. Given the existence of the large number of sequence variants in native Africans,[2,3] it would be prudent to perform WES on carriers of *PDE6B* and *CERKL* variants to identify causal IRD mutation(s).

***PRPF3*: c.1480A>G; p.(Thr494Ala)**

Human (Homo sapiens)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Chimpanzee (Pan troglodytes)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Cow (Bos taurus)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Dog (Canis lupus familiaris)	GAAGCTGTTCAAGACCCC**A**CGAAGGTAGAAGCCCA
Zebrafish (Danio rerio)	GAAGCTGTGCAGGACCCC**A**CTAAAGTGGAGGCCCA
Chicken (Gallus gallus)	GAAGCTGTTCAGGACCCA**A**CAAAAGTTGAAGCTCA
Mouse (Mus musculus)	GAAGCAGTTCAAGACCCT**A**CAAAGGTAGAAGCCCA

***RHO*: c.154T>G; p.(Phe52Val)**

Human (Homo sapiens)	CTGCTGATCGTGCTGGGC**T**TCCCCATCAACTTCCTC
Chimpanzee (Pan troglodytes)	CTGCTGATCGTGCTGGGC**T**TCCCCATCAACTTCCTC
Cow (Bos taurus)	CTGCTGATCATGCTTGGC**T**TCCCCATCAACTTCCTC
Dog (Canis lupus familiaris)	CTGCTGATCGTGCTCGGC**T**TCCCCATCAACTTCCTC
Zebrafish (Danio rerio)	TTCCTCATCATCACCGGC**T**TCCCCGTCAACTTCCTC
Chicken (Gallus gallus)	ATGCTGATCCTGCTCGGC**T**TCCCCGTCAACTTCCTC
Mouse (Mus musculus)	CTGCTCATCGTGCTGGGC**T**TCCCCATCAACTTCCTC

***PRPF31*: c.698-1G>A; p.(?)**

Human (Homo sapiens)	CTCCCTCCCACCGCA**G**GTGGGCCGGCGGCCTGACC
Chimpanzee (Pan troglodytes)	CTCCCTCCCACCGCA**G**GTGTGGCCGGCGGCCTGACG
Cow (Bos taurus)	----------CA**G**GGGTGGCCGGCGGCCTGACC
Dog (Canis lupus familiaris)	CTCCTTCCTGTTGCA**G**GGGTGGCTGGGGGCCTGACC
Zebrafish (Danio rerio)	TTTCCTTTTATTTCA**G**GTGTTGCTGGTGGTCTGACT
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	----------GCA**G**GTGTGGCTGGAGGGCTCACC

***ABCA4*: c.4832delC; p.(Thr1611MetfsTer51)**

Human (Homo sapiens)	AAACATCTAG AAA**C**TGAAGA CAACATTAAG GTACTT
Chimpanzee (Pan troglodytes)	AAACATCTAG AAA**C**TGAAGA CAACATTAAG GTACTT
Cow (Bos taurus)	AAACAACTAG AAA**C**TGAAGA CAATATTAAG GTATTG
Dog (Canis lupus familiaris)	AAACATCTAG AAA**C**TGAAGA CAACATTAAG GTATTG
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	AAACATCTTG AAA**C**CACAGA CAACATTAAG GTAC--

***ABCA4*: c.1043T>G; p.(Phe348Cys)**

Human (Homo sapiens)	TATAAGGCCT**T**TCTGGGGAT TGACTCCACA AGGAAG
Chimpanzee (Pan troglodytes)	TATAAGGCCT**T**TCTGGGGAT TGACTCCACA AGGAAG
Cow (Bos taurus)	TATAAGGCCT**T**CTAGGGAT CGACTCCACA AGGAAA
Dog (Canis lupus familiaris)	TATAAGGCCT**T**CTTAGGGAT TGACTCCACA AGGAAG
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	TACAAAGCTT**T**CCTGGGGAT TGATTCCACA AAGAAA
Mouse (Mus musculus)	TATAAAGCCT**T**CCTGGGGAT TGATTCCACA AGGAAA

***PDE6B*: c.1860delC; p.(His620GlnfsTer23)**

Human (Homo sapiens)	TGGCTAAGCTCCA**C**GGCTCC TCGATTTTGG AGCGGC
Chimpanzee (Pan troglodytes)	no homologue
Cow (Bos taurus)	TAGCCAAGCTCCA**C**GGCTCC TCGATTTTGG AGCGAC
Dog (Canis lupus familiaris)	TGGCCAAGCTCCA**C**GGCTCC TCCATCCTGG AGCGGC
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	TAGCCAAGTTACATGGCTCCTCAATTCTGGAAAGGC

***CERKL*: c.365T>G; p.(Leu122Arg)**

Human (Homo sapiens)	TTTATTAGGTATCACAC**T**CTTCATCTGCTTGAAAAA
Chimpanzee (Pan troglodytes)	TTTATTAGGTATCACAC**T**CTTCATCTGCTTGAAAAA
Cow (Bos taurus)	TTTATTAGGAATCACTC**T**CTTCATATGTTTGAAAAC
Dog (Canis lupus familiaris)	TTTATTGGGTATCACAC**T**CTTCATTTGTTTGAAAAA
Zebrafish (Danio rerio)	no homologue
Chicken (Gallus gallus)	no homologue
Mouse (Mus musculus)	CTGCTGGGCATAACGC**T**TTTCATCTGCTTGAAGGA

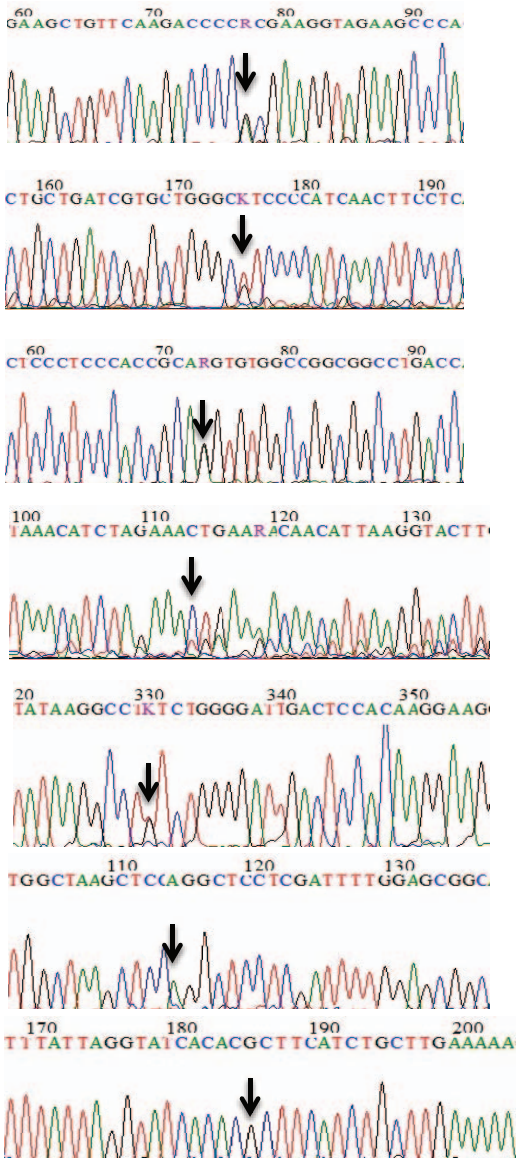

Figure 2. Sequence alignments across vertebrate species demonstrating nucleotide conservation of each identified variant (highlighted in *red*). *Right*: the corresponding Sanger sequencing electropherogram. *Arrow* indicates the position of the mutation.

Our study shows that genetic investigations of the SA indigenous population present considerable challenges and unique opportunities in human disease gene discovery.[49] Africans have smaller haplotype blocks and low levels of linkage disequilibrium compared to non-African populations, as well as evidence of genetic admixture, leading to unique diversity.[3,4] Whole-exome sequencing of RP families in the United States has yielded a greater number of novel variants (both single nucleotide variants and small indels) in the families of African ancestry compared with families of European ancestry.[50] In this study, the number of variants novel to the National Center for Biotechnology Information Short Genetic Variations database (dbSNP) was reportedly >6-fold larger in a family of African American descent ($n > 2500$) than in Caucasian U.S. families ($n \sim 400$). Given that genome-wide ancestry estimates show an average proportion of only ~73% African ancestry in African Americans,[51] the exomes of indigenous Africans are expected to yield even more novel variants. Therefore, inclusion of African populations in genomics research should facilitate the discovery of genetic defects associated with human disease.[52]

This study employs the first next generation sequencing (NGS)-based approach in an indigenous SA cohort as an opportunity for improved understanding of the genetic architecture of IRDs. We have shown that success of diagnosis is enhanced considerably using WES, and have identified important genes and novel variants for genetic counseling for IRD patients. Our study provides valuable insight into the etiology of IRD in SA, and contributes toward more comprehensive understanding of this heterogeneous group of disorders by cataloguing novel causative variants.

Acknowledgments

The authors thank patients and family members for participation and Linn Gieser and Ash-Police Reddy for technical assistance with WES. This study made use of data generated by the African Genomes Variation Project, for the African Partnership for Chronic

Disease Research. A full list of the investigators and funders who contributed to the generation or collation of the data is available from www.apcdr.org.

Research in South Africa was funded by Retina South Africa and the Medical Research Council of South Africa. WES and data analysis were supported by the Intramural Research Program (EY000546) of the National Eye Institute and utilized computational resources of the National Institutes of Health High-Performance Computing Biowulf cluster (https://hpc.nih.gov).

Disclosure: **L. Roberts**, None; **R. Ratnapriya**, None; **M. du Plessis**, None; **V. Chaitankar**, None; **R.S. Ramesar**, None; **A. Swaroop**, None

References

1. Campbell MC, Tishkoff SA. The evolution of human genetic and phenotypic variation in Africa. *Curr Biol*. 2010;20:R166–R173.
2. May A, Hazelhurst S, Li Y, et al. Genetic diversity in black South Africans from Soweto. *BMC Genomics*. 2013;14:644.
3. Schuster SC, Miller W, Ratan A, et al. Complete Khoisan and Bantu genomes from southern Africa. *Nature*. 2010;463:943–947.
4. Chimusa ER, Meintjies A, Tchanga M, et al. A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. *PLoS Genet*. 2015;11:e1005052.
5. Li S, Schlebusch C, Jakobsson M. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proc Biol Sci*. 2014;281.
6. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res*. 2010;29:335–375.
7. Wright AF, Chakarova CF, Abd El-Aziz MM, Bhattacharya SS. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat Rev Genet*. 2010;11:273–284.
8. Swaroop A, Kim D, Forrest D. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat Rev Neurosci*. 2010;11:563–576.
9. Veleri S, Lazar CH, Chang B, Sieving PA, Banin E, Swaroop A. Biology and therapy of inherited retinal degenerative disease: insights from mouse models. *Dis Model Mech*. 2015;8:109–129.
10. Anasagasti A, Irigoyen C, Barandika O, de Munain AL, Ruiz-Ederra J. Current mutation discovery approaches in retinitis pigmentosa. *Vis Res*. 2012;75:117–129.
11. Nishiguchi KM, Rivolta C. Genes associated with retinitis pigmentosa and allied diseases are frequently mutated in the general population. *PLoS One*. 2012;7:e41902.
12. Daiger SP, Bowne SJ, Sullivan LS. Genes and mutations causing autosomal dominant retinitis pigmentosa. *Cold Spring Harb Perspect Med*. 2015;5:a017129.
13. Greenberg J, Bartmann L, Ramesar R, Beighton P. Retinitis pigmentosa in southern Africa. *Clin Genet*. 1993;44:232–235.
14. Greenberg J, Roberts L, Bruwer Z, Schoeman M, Loggenberg K, Loubser F. Delivery of an ophthalmic genetic service in South Africa. *S A Ophthalmol J*. 2010;5:14–19.
15. Dalkara D, Goureau O, Marazova K, Sahel JA. Let there be light: gene and cell therapy for blindness. *Hum Gene Ther*. 2016;27:134–147.
16. Roberts L, Goliath R, Rebello G, et al. Inherited retinal disorders in South Africa and the clinical impact of evolving technologies. *S Afr Med J*. 2016;106:10988.
17. Roberts L, Bartmann L, Ramesar R, Greenberg J. Novel variants in the hotspot region of RP1 in South African patients with retinitis pigmentosa. *Mol Vis*. 2006;12:177–183.
18. Roberts L, Ramesar R, Greenberg J. Low frequency of rhodopsin mutations in South African patients with autosomal dominant retinitis pigmentosa. *Clin Genet*. 2000;58:77–78.
19. Roberts L, Rebello G, Greenberg J, Ramesar R. Great expectations: RPE65 mutations in South Africa. In: Baert M, Peeters C, eds. *Retinitis Pigmentosa: Causes Diagnosis and Treatment*. New York, NY: Nova Science Publishers; 2010:89–110.
20. Greenberg J, Roberts L, Ramesar R. Unusual frequencies of rhodopsin mutations and polymorphisms in South African patients with retinitis tigmentosa. In: Anderson RE, LaVail MM, Hollyfield JG, eds. *New Insights into Retinal Degenerative Diseases. A Book on the Proceedings of the IXth International Symposium on Retinal Degeneration*. New York: Kluwer Academic/Plenum Publishers; 2002:329–331.
21. Ratnapriya R, Swaroop A. Genetic architecture of retinal and macular degenerative diseases: the promise and challenges of next-generation sequencing. *Genome Med*. 2013;5:84.
22. Lazar CH, Mutsuddi M, Kimchi A, et al. Whole exome sequencing reveals GUCY2D as a major gene associated with cone and cone-rod dystrophy in Israel. *Invest Ophthalmol Vis Sci*. 2015;56:420–430.
23. Thompson DA, Ali RR, Banin E, et al. Advancing therapeutic strategies for inherited retinal degeneration: recommendations from the Monaciano Symposium. *Invest Ophthalmol Vis Sci*. 2015;56:918–931.
24. Chaitankar V, Karakulah G, Ratnapriya R, Giuste FO, Brooks MJ, Swaroop A. Next generation sequencing technology and genomewide data analysis: perspectives for retinal research [published online ahead of print June 10, 2016]. *Prog Retin Eye Res*. doi: 10.1016/j.preteyeres.2016.06.001.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–2120.
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.
27. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–498.
28. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303.
29. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
30. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
31. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat*. 2013;34:E2393–E2402.
32. Cunningham F, Amode MR, Barrell D, et al. Ensembl 2015. *Nucleic Acids Res*. 2015;43:D662–D669.
33. Gurdasani D, Carstensen T, Tekola-Ayele F, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 2015;517:327–332.
34. Danciger M, Blaney J, Gao YQ, et al. Mutations in the PDE6B gene in autosomal recessive retinitis pigmentosa. *Genomics*. 1995;30:1–7.
35. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37:e67.
36. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics

and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424.

37. Audo I, Bujakowska KM, Leveillard T, et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. *Orphanet J Rare Dis*. 2012;7:8.
38. Booij JC, Bakker A, Kulumbetova J, et al. Simultaneous mutation detection in 90 retinal disease genes in multiple patients using a custom-designed 300-kb retinal resequencing chip. *Ophthalmology*. 2011;118:160–167.
39. Ge Z, Bowles K, Goetz K, et al. NGS-based Molecular diagnosis of 105 eyeGENE((R)) probands with Retinitis Pigmentosa. *Sci Rep*. 2015;5:18287.
40. O'Sullivan J, Mullaney BG, Bhaskar SS, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *J Med Genet*. 2012;49:322–326.
41. Wang X, Wang H, Sun V, et al. Comprehensive molecular diagnosis of 179 Leber congenital amaurosis and juvenile retinitis pigmentosa patients by targeted next generation sequencing. *J Med Genet*. 2013;50:674–688.
42. Consugar MB, Navarro-Gomez D, Place EM, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection than exome sequencing. *Genet Med*. 2015;17:253–261.
43. Corton M, Nishiguchi KM, Avila-Fernandez A, et al. Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. *PLoS One*. 2013;8:e65574.
44. Abu-Safieh L, Alrashed M, Anazi S, et al. Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res*. 2013;23:236–247.
45. Xu Y, Guan L, Shen T, et al. Mutations of 60 known causative genes in 157 families with retinitis pigmentosa based on exome sequencing. *Hum Genet*. 2014;133:1255–1271.
46. Jinda W, Taylor TD, Suzuki Y, et al. Whole exome sequencing in Thai patients with retinitis pigmentosa reveals novel mutations in six genes. *Invest Ophthalmol Vis Sci*. 2014;55:2259–2268.
47. Beryozkin A, Shevah E, Kimchi A, et al. Whole exome sequencing reveals mutations in known retinal disease genes in 33 out of 68 Israeli families with inherited retinopathies. *Sci Rep*. 2015;5:13187.
48. Manes G, Cheguru P, Majumder A, et al. A truncated form of rod photoreceptor PDE6 beta-subunit causes autosomal dominant congenital stationary night blindness by interfering with the inhibitory activity of the gamma-subunit. *PLoS One*. 2014;9:e95768.
49. Ramsay M, Tiemessen CT, Choudhury A, Soodyall H. Africa: the next frontier for human disease gene discovery? *Hum Mol Genet*. 2011;20:R214–R220.
50. Koboldt DC, Larson DE, Sullivan LS, et al. Exome-based mapping and variant prioritization for inherited Mendelian disorders. *Am J Hum Genet*. 2014;94:373–384.
51. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet*. 2015;96:37–53.
52. Sirugo G, Hennig BJ, Adeyemo AA, et al. Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. *Hum Genet*. 2008;123:557–598.

Investigative Ophthalmology & Visual Science